STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 166245

TO: Jeffrey Fredman
Location: rem/2C89/2C18
Art Unit: 1637
Friday, October 14, 2005

Case Serial Number: 09/744097

From: Edward Hart
Location: Biotech-Chem Library
REM-1A55
Phone: 571-272-2512

edward.hart@uspto.gov

Search Notes

Examiner Fredman,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib 168245 mf

From: Fredman, Jeffrey
Sent: Tuesday, October 11, 2005 10:23 AM
To: STIC-Biotech/ChemLib
Subject: 09/744,097

Please search SEQ ID NO: 76 in nucleic acid databases.

Thanks,

Jeffrey Fredman
Art Unit 1637
Remsen Building 2C89
(571)272-0742

2C18

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up: 10/12/05
Date completed: 10/14/05
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#__1__ AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:____
Inventor:_____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM: OBH
WWW/Internet:_____
Other (Specify):_____

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 13, 2005, 17:46:38 ; Search time 1479 Seconds
(without alignments)
884.578 Million cell updates/sec

Title: US-09-744-097A-76
Perfect score: 27
Sequence: 1 gtagcctagctacccctaggtctaggc 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	20.6	76.3	135412	2	AC148279	AC148279 Sorex ara
	2	20.6	76.3	156661	2	AC148358	AC148358 Sorex ara
	3	20.6	76.3	278310	2	AC127955	AC127955 Rattus no
c	4	20.6	76.3	285193	2	AC111242	AC111242 Rattus no
c	5	20.6	76.3	293962	2	AC112303	AC112303 Rattus no
	6	20.4	75.6	188946	10	AC127236	AC127236 Mus muscu
c	7	20.4	75.6	194613	2	AC116089	AC116089 Rattus no
c	8	20.2	74.8	124104	8	AC134931	AC134931 Oryza sat
	9	19.8	73.3	119743	2	AC134515	AC134515 Rattus no
	10	19.8	73.3	211465	2	AC103070	AC103070 Rattus no
	11	19.8	73.3	212559	2	AC095896	AC095896 Rattus no
	12	19.6	72.6	857	4	SSAPOA11	X69478 S.scrofa ge
	13	19.6	72.6	5551	10	AK173314	AK173314 Mus muscu
c	14	19.6	72.6	207223	10	AL732521	AL732521 Mouse DNA
	15	19.6	72.6	212481	10	AC126959	AC126959 Rattus no
	16	19.6	72.6	252851	2	AC135094	AC135094 Rattus no
c	17	19.4	71.9	276137	2	AC131170	AC131170 Rattus no
	18	19.2	71.1	91977	10	AL732588	AL732588 Mouse DNA
	19	19.2	71.1	116926	10	AL808012	AL808012 Mouse DNA
	20	19.2	71.1	147467	2	AC118573	AC118573 Lemur cat
c	21	19.2	71.1	186449	2	AC147134	AC147134 Mus muscu
	22	19.2	71.1	188892	2	AC091327	AC091327 Mus muscu
c	23	19.2	71.1	237378	2	AC087038	AC087038 Mus muscu
	24	19	70.4	110000	1	AE017180_27	Continuation (28 o
	25	19	70.4	150222	2	AC120016	AC120016 Mus muscu
	26	19	70.4	164530	10	AL596104	AL596104 Mouse DNA
	27	19	70.4	187514	2	AC109240	AC109240 Mus muscu
c	28	19	70.4	203946	2	AC069465	AC069465 Mus muscu
	29	19	70.4	233852	2	AC108549	AC108549 Rattus no
c	30	19	70.4	235241	10	AL928893	AL928893 Mouse DNA
c	31	19	70.4	240264	2	AC107434	AC107434 Rattus no
	32	19	70.4	245134	2	AC126639	AC126639 Rattus no
	33	19	70.4	248917	2	AC111455	AC111455 Rattus no
c	34	19	70.4	250579	2	AC126843	AC126843 Rattus no
	35	19	70.4	329223	2	AC109504	AC109504 Mus muscu
c	36	18.8	69.6	81368	2	AC094244_4	Continuation (5 of
c	37	18.8	69.6	103428	2	AC096436_6	Continuation (7 of
c	38	18.8	69.6	104113	10	AL928640	AL928640 Mouse DNA
c	39	18.8	69.6	110000	2	AC102028_3	Continuation (4 of
	40	18.8	69.6	142937	2	AC139916	AC139916 Rattus no
c	41	18.8	69.6	148320	10	AL669849	AL669849 Mouse DNA
c	42	18.8	69.6	154685	10	AL606511	AL606511 Mouse DNA
	43	18.8	69.6	157088	10	AC127292	AC127292 Mus muscu
c	44	18.8	69.6	171936	10	AL929035	AL929035 Mouse DNA
c	45	18.8	69.6	173598	10	AL672194	AL672194 Mouse DNA

ALIGNMENTS

RESULT 1
AC148279
LOCUS AC148279 135412 bp DNA linear HTG 15-APR-2004
DEFINITION Sorex araneus clone SA_Ba-546N4, WORKING DRAFT SEQUENCE, 7 ordered pieces.
ACCESSION AC148279
VERSION AC148279.2 GI:46391177
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Sorex araneus (European shrew)
ORGANISM Sorex araneus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Insectivora; Soricidae; Soricinae; Sorex.
REFERENCE 1 (bases 1 to 135412)
AUTHORS Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B., Coleman,H., Daki,N., Engle,J., Granite,S., Guan,X., Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B., Idol,J.R., Jones,C., Karlins,E., Kim,H., Kwong,P., Laric,P., Larson,S., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J., Mullikin,J.C., Paguirigan,C., Portnoy,M.E., Prasad,A., Puri,O., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
TITLE NISC Comparative Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 135412)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2004) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE 3 (bases 1 to 135412)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (15-APR-2004) NIH Intramural Sequencing Center, 8717 Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT On Apr 15, 2004 this sequence version replaced gi:42627935.
---------- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov

```
            ---------- Project Information
                 Center project name: gbi
                 Center clone name: 546N04

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 134185 bases at least Q40
                 Consensus quality: 134570 bases at least Q30
                 Consensus quality: 134696 bases at least Q20
                 Insert size: 118000; agarose-fp
                 Insert size: 134812; sum-of-contigs
                 Quality coverage: 11.34x in Q20 bases; agarose-fp
                 Quality coverage: 9.93x in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    28824: contig of 28824 bp in length
            *    28825    28924: gap of unknown length
            *    28925    35906: contig of 6982 bp in length
            *    35907    36006: gap of unknown length
            *    36007    61654: contig of 25648 bp in length
            *    61655    61754: gap of unknown length
            *    61755    68663: contig of 6909 bp in length
            *    68664    68763: gap of unknown length
            *    68764   126764: contig of 58001 bp in length
            *   126765   126864: gap of unknown length
            *   126865   133235: contig of 6371 bp in length
            *   133236   133335: gap of unknown length
            *   133336   135412: contig of 2077 bp in length.
FEATURES             Location/Qualifiers
     source          1. .135412
                     /organism="Sorex araneus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:42254"
                     /clone="SA_Ba-546N4"
                     /clone_lib="SA_Ba"
     misc_feature    1. .60485
                     /note="clone overlaps with GenBank Accession Number
                     AC148351 clone SA_Ba-157G15 (center project name gbf)"
     misc_feature    1. .28824
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    28925. .35906
                     /note="assembly_fragment"
     misc_feature    36007. .61654
                     /note="assembly_fragment"
     misc_feature    61755. .68663
                     /note="assembly_fragment"
     misc_feature    68764. .126764
                     /note="assembly_fragment"
     misc_feature    69707. .135412
                     /note="clone overlaps with GenBank Accession Number
                     AC148358 clone SA_Ba-621C10 (center project name gbh)"
     misc_feature    126865. .133235
                     /note="assembly_fragment"
     misc_feature    133336. .135412
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             76.3%;  Score 20.6;  DB 2;  Length 135412;
  Best Local Similarity   85.2%;  Pred. No. 20;
  Matches   23;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              ||| |||| |||||||||| | |||||
Db     109081 GTACCCTAACTACCCCTAGATTTAGGC 109107


RESULT 2
AC148358
LOCUS       AC148358              156661 bp    DNA     linear   HTG 27-MAR-2004
DEFINITION  Sorex araneus clone SA_Ba-621C10, WORKING DRAFT SEQUENCE, 11
            ordered pieces.
ACCESSION   AC148358
VERSION     AC148358.2  GI:45773965
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Sorex araneus (European shrew)
  ORGANISM  Sorex araneus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Insectivora; Soricidae; Soricinae; Sorex.
REFERENCE   1  (bases 1 to 156661)
  AUTHORS   Antonellis,A., Ayele,K., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Chu,G., Coleman,B.,
            Coleman,H., Daki,N., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P., Hurle,B.,
            Idol,J.R., Jones,C., Karlins,E., Kim,H., Kwong,P., Laric,P.,
            Larson,S., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Mullikin,J.C., Paguirigan,C., Portnoy,M.E., Prasad,A., Puri,O.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V., Vogt,J.L.,
            Wetherby,K.D., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156661)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-FEB-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 156661)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2004) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Mar 27, 2004 this sequence version replaced gi:42821792.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: gbh
                 Center clone name: 621C10

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
```

```
          coverage in Q20 bases and has been reviewed to rule out
          gross misassemblies, the low-quality ends of sequence
          contigs have been trimmed away, and each base is associated
          with a Phrap-derived quality score.
          ---------- Summary Statistics
               Sequencing vector: plasmid; n/a; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.990319
               Consensus quality: 154585 bases at least Q40
               Consensus quality: 155145 bases at least Q30
               Consensus quality: 155506 bases at least Q20
               Insert size: 107000; agarose-fp
               Insert size: 155661; sum-of-contigs
               Quality coverage: 12.02x in Q20 bases; agarose-fp
               Quality coverage: 8.26x in Q20 bases; sum-of-contigs
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *        1    41408: contig of 41408 bp in length
          *    41409    41508: gap of unknown length
          *    41509    43634: contig of 2126 bp in length
          *    43635    43734: gap of unknown length
          *    43735    57261: contig of 13527 bp in length
          *    57262    57361: gap of unknown length
          *    57362    63861: contig of 6500 bp in length
          *    63862    63961: gap of unknown length
          *    63962    97769: contig of 33808 bp in length
          *    97770    97869: gap of unknown length
          *    97870   108599: contig of 10730 bp in length
          *   108600   108699: gap of unknown length
          *   108700   129177: contig of 20478 bp in length
          *   129178   129277: gap of unknown length
          *   129278   134690: contig of 5413 bp in length
          *   134691   134790: gap of unknown length
          *   134791   141931: contig of 7141 bp in length
          *   141932   142031: gap of unknown length
          *   142032   151911: contig of 9880 bp in length
          *   151912   152011: gap of unknown length
          *   152012   156661: contig of 4650 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156661
                     /organism="Sorex araneus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:42254"
                     /clone="SA_Ba-621C10"
                     /clone_lib="SA_Ba"
     misc_feature    1. .67579
                     /note="clone overlaps with GenBank Accession Number
                     AC148279 clone SA_Ba-546N4 (center project name gbi)"
     misc_feature    1. .41408
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    41509. .43634
                     /note="assembly_fragment"
     misc_feature    43735. .57261
                     /note="assembly_fragment"
     misc_feature    57362. .63861
                     /note="assembly_fragment"
     misc_feature    63962. .97769
                     /note="assembly_fragment"
     misc_feature    97870. .108599
                     /note="assembly_fragment"
     misc_feature    108700. .129177
                     /note="assembly_fragment"
     misc_feature    129278. .134690
                     /note="assembly_fragment"
     misc_feature    129878. .156661
                     /note="clone overlaps with GenBank Accession Number
                     AC148355 clone SA_Ba-408018 (center project name gga)"
     misc_feature    134791. .141931
                     /note="assembly_fragment"
     misc_feature    142032. .151911
                     /note="assembly_fragment"
     misc_feature    152012. .156661
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             76.3%;  Score 20.6;  DB 2;  Length 156661;
  Best Local Similarity   85.2%;  Pred. No. 20;
  Matches   23;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              ||| |||| |||||||||| | |||||
Db      39469 GTACCCTAACTACCCCTAGATTTAGGC 39495


RESULT 3
AC127955
LOCUS       AC127955              278310 bp    DNA     linear   HTG 19-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-270L18, *** SEQUENCING IN PROGRESS
            ***, 6 unordered pieces.
ACCESSION   AC127955
VERSION     AC127955.2  GI:23195969
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 278310)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
```

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 278310)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 278310)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Sep 19, 2002 this sequence version replaced gi:21908490.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequening reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: KABR
Center clone name: CH230-270L18
---------- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 159405 bases at least Q40
Consensus quality: 163261 bases at least Q30
Consensus quality: 165346 bases at least Q20
Estimated insert size: 206482; sum-of-contigs estimation
Quality coverage: 3x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 197344: contig of 197344 bp in length
* 197345 197444: gap of unknown length
* 197445 266948: contig of 69504 bp in length
* 266949 267048: gap of unknown length
* 267049 269714: contig of 2666 bp in length
* 269715 269814: gap of unknown length
* 269815 271799: contig of 1985 bp in length
* 271800 271899: gap of unknown length
* 271900 273966: contig of 2067 bp in length
* 273967 274066: gap of unknown length
* 274067 278310: contig of 4244 bp in length.
FEATURES Location/Qualifiers

source 1. .278310
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-270L18"
misc_feature 1. .2376
/note="wgs_contig"
misc_feature 153418. .156139
/note="wgs_contig"
misc_feature 211774. .212782
/note="wgs_contig"
ORIGIN

Query Match 76.3%; Score 20.6; DB 2; Length 278310;
Best Local Similarity 85.2%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||| ||||||| |||||| ||| |
Db     248352 GTAGCCAAGCTACCTCTAGGTGTAGAC 248378
```

RESULT 4
AC111242/c
LOCUS AC111242 285193 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-232H4, *** SEQUENCING IN PROGRESS ***, 11 unordered pieces.
ACCESSION AC111242
VERSION AC111242.4 GI:30578456
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 285193)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 285193)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 285193)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 13, 2003 this sequence version replaced gi:23604121.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GLWE
Center clone name: CH230-232H4
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 242831 bases at least Q40
Consensus quality: 249455 bases at least Q30
Consensus quality: 253771 bases at least Q20
Estimated insert size: 272024; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 5183: contig of 5183 bp in length
* 5184 5283: gap of unknown length
* 5284 17822: contig of 12539 bp in length
* 17823 17922: gap of unknown length
* 17923 116644: contig of 98722 bp in length
* 116645 116744: gap of unknown length
* 116745 157368: contig of 40624 bp in length
* 157369 157468: gap of unknown length
* 157469 247903: contig of 90435 bp in length
* 247904 248003: gap of unknown length
* 248004 261891: contig of 13888 bp in length
* 261892 261991: gap of unknown length
* 261992 276510: contig of 14519 bp in length
* 276511 276610: gap of unknown length
* 276611 277878: contig of 1268 bp in length
* 277879 277978: gap of unknown length
* 277979 279325: contig of 1347 bp in length
* 279326 279425: gap of unknown length
* 279426 282175: contig of 2750 bp in length
* 282176 282275: gap of unknown length
* 282276 285193: contig of 2918 bp in length.

FEATURES Location/Qualifiers
source 1. .285193
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-232H4"
misc_feature 1. .2558
/note="wgs_end_extension
clone_end:T7"
misc_feature complement(7161. .8031)
/note="clone_boundary
clone_end:T7
site:EcoRI
end_sequence:BZ092926"
misc_feature 16581. .17822
/note="wgs_contig"
misc_feature 30287. .34910
/note="wgs_contig"
misc_feature 34961. .39614
/note="wgs_contig"
misc_feature 40916. .42387
/note="wgs_contig"
misc_feature 58872. .59764
/note="clone_boundary
clone_end:Sp6
site:EcoRI
end_sequence:BZ092929"
misc_feature 113101. .116644
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 136932. .138724
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 155552. .157368
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 257732. .259507
/note="wgs_end_extension
clone_end:Sp6"
misc_feature 268633. .274031
/note="wgs_end_extension
clone_end:Sp6"

ORIGIN

Query Match 76.3%; Score 20.6; DB 2; Length 285193;
Best Local Similarity 85.2%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
|||||| ||||||| |||||| ||| |
Db 114184 GTAGCCAAGCTACCTCTAGGTGTAGAC 114158

RESULT 5
AC112303/c
LOCUS AC112303 293962 bp DNA linear HTG 15-NOV-2002
DEFINITION Rattus norvegicus clone CH230-208P20, *** SEQUENCING IN PROGRESS
***, 8 unordered pieces.
ACCESSION AC112303

```
VERSION     AC112303.4  GI:24635626
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 293962)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 293962)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 293962)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 6, 2002 this sequence version replaced gi:23603916.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMYH
                Center clone name: CH230-208P20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 206616 bases at least Q40
                Consensus quality: 209134 bases at least Q30
                Consensus quality: 211305 bases at least Q20
                Estimated insert size: 210335; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    27340: contig of 27340 bp in length
            *    27341    27440: gap of unknown length
            *    27441    51601: contig of 24161 bp in length
            *    51602    51701: gap of unknown length
            *    51702   268441: contig of 216740 bp in length
            *   268442   268541: gap of unknown length
            *   268542   270645: contig of 2104 bp in length
            *   270646   270745: gap of unknown length
            *   270746   272238: contig of 1493 bp in length
            *   272239   272338: gap of unknown length
            *   272339   275354: contig of 3016 bp in length
            *   275355   275454: gap of unknown length
            *   275455   281008: contig of 5554 bp in length
            *   281009   281108: gap of unknown length
            *   281109   293962: contig of 12854 bp in length.
FEATURES             Location/Qualifiers
     source          1. .293962
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-208P20"
     misc_feature    6200. .7062
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:RWBKP94TVB"
     misc_feature    51702. .53500
                     /note="wgs_contig"
     misc_feature    53551. .54607
                     /note="wgs_contig"
     misc_feature    56203. .57503
                     /note="wgs_contig"
     misc_feature    174983. .176056
                     /note="wgs_contig"
     misc_feature    208600. .209401
                     /note="clone_boundary
```

```
                     clone_end:T7
                     site:
                     end_sequence:RWBKP94TJB"
     misc_feature    264125. .265539
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    267070. .268441
                     /note="wgs_end_extension
                     clone_end:T7"
ORIGIN

  Query Match             76.3%;  Score 20.6;  DB 2;  Length 293962;
  Best Local Similarity   85.2%;  Pred. No. 19;
  Matches   23;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              ||||||  ||||||| ||||||  ||| |
Db     256250 GTAGCCAAGCTACCTCTAGGTGTAGAC 256224


RESULT 6
AC127236
LOCUS       AC127236               188946 bp    DNA     linear   ROD 27-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-351L1 from chromosome 18, complete
            sequence.
ACCESSION   AC127236
VERSION     AC127236.3  GI:37361086
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 188946)
  AUTHORS   Swearengen-Shahid,S.
  TITLE     The sequence of Mus musculus BAC clone RP24-351L1
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 188946)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 188946)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 188946)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 188946)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 188946)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Oct 2, 2003 this sequence version replaced gi:23308121.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BB0351L01
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-24 BAC Library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone. This clone is
            overlapped by AC108434.
FEATURES             Location/Qualifiers
     source          1. .188946
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="18"
                     /map="18"
                     /clone="RP24-351L1"
                     /clone_lib="RPCI-24"
     repeat_region   773. .874
                     /rpt_family="Alu"
     repeat_region   920. .1078
                     /rpt_family="Alu"
     repeat_region   1428. .1643
                     /rpt_family="B2"
     repeat_region   1675. .1845
                     /rpt_family="Alu"
     repeat_region   2684. .2873
                     /rpt_family="B2"
     repeat_region   3164. .3303
                     /rpt_family="Alu"
     repeat_region   3577. .3670
                     /rpt_family="Alu"
     repeat_region   3688. .3892
                     /rpt_family="B2"
     repeat_region   5876. .6150
                     /rpt_family="B4"
     repeat_region   6252. .6351
                     /rpt_family="Alu"
     repeat_region   6452. .6680
                     /rpt_family="B4"
     repeat_region   6545. .6681
                     /rpt_family="Alu"
     repeat_region   6698. .6951
                     /rpt_family="B4"
     repeat_region   7324. .7516
                     /rpt_family="B2"
     repeat_region   10777. .10894
                     /rpt_family="B4"
     repeat_region   11023. .11190
                     /rpt_family="B4"
     repeat_region   11031. .11227
                     /rpt_family="B2"
     repeat_region   11303. .11433
                     /rpt_family="Alu"
     repeat_region   11547. .11724
                     /rpt_family="B2"
     repeat_region   11736. .12221
```

```
                     /rpt_family="L1"
     repeat_region   12388. .12444
                     /rpt_family="B4"
     repeat_region   12393. .12555
                     /rpt_family="Alu"
     repeat_region   12919. .13263
                     /rpt_family="ERVK"
     repeat_region   13690. .13828
                     /rpt_family="Alu"
     repeat_region   13846. .14036
                     /rpt_family="B2"
     repeat_region   14431. .14567
                     /rpt_family="Alu"
     repeat_region   14704. .14812
                     /rpt_family="Alu"
     repeat_region   16470. .16651
                     /rpt_family="B2"
     repeat_region   16751. .16836
                     /rpt_family="B4"
     repeat_region   16907. .16994
                     /rpt_family="Alu"
     repeat_region   17163. .17302
                     /rpt_family="B4"
     repeat_region   17422. .17450
                     /rpt_family="MaLR"
     repeat_region   17451. .17642
                     /rpt_family="B2"
     tRNA            complement(17566. .17637)
                     /product="tRNA-Ser"
                     /note="Likely pseudogene (HMM Sc=32.99 / Sec struct
                     Sc=-11.32)"
     repeat_region   17643. .17968
                     /rpt_family="MaLR"
     repeat_region   18665. .18753
                     /rpt_family="BC1_MM"
     repeat_region   18685. .18756
                     /rpt_family="B4"
     repeat_region   18865. .19010
                     /rpt_family="Alu"
     repeat_region   19368. .19574
                     /rpt_family="B4"
     repeat_region   20046. .20209
                     /rpt_family="B4"
     repeat_region   20485. .20641
                     /rpt_family="Alu"
     repeat_region   21092. .21114
                     /rpt_family="MaLR"
     repeat_region   21115. .21504
                     /rpt_family="ERVK"
     repeat_region   21505. .21858
                     /rpt_family="MaLR"
     repeat_region   21868. .21995
                     /rpt_family="Alu"
     repeat_region   22114. .22212
                     /rpt_family="Alu"
     repeat_region   22862. .23049
                     /rpt_family="B2"
     repeat_region   23410. .23556
                     /rpt_family="Alu"
     repeat_region   23563. .23703
                     /rpt_family="Alu"
     repeat_region   24073. .24298
                     /rpt_family="L1"
     repeat_region   24299. .24497
                     /rpt_family="B2"
     repeat_region   24498. .24584
                     /rpt_family="L1"
     repeat_region   24585. .24732
                     /rpt_family="Alu"
     repeat_region   24908. .25055
                     /rpt_family="Alu"
     repeat_region   25676. .25782
                     /rpt_family="Alu"
     repeat_region   25783. .25991
                     /rpt_family="MaLR"
     repeat_region   26004. .26094

  Query Match           75.6%;  Score 20.4;  DB 10;  Length 188946;
  Best Local Similarity   95.5%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 CTAGCTACCCCTAGGTCTAGGC 27
              ||| ||||||||||||||||||
Db     132397 CTAACTACCCCTAGGTCTAGGC 132418


RESULT 7
AC116089/c
LOCUS       AC116089              194613 bp    DNA     linear   HTG 20-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-344O11, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC116089
VERSION     AC116089.6  GI:25138296
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 194613)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
```

```
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194613)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 194613)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 20, 2002 this sequence version replaced gi:23194943.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GTCO
                Center clone name: CH230-344O11
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 178038 bases at least Q40
                Consensus quality: 179671 bases at least Q30
                Consensus quality: 180674 bases at least Q20
                Estimated insert size: 180448; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   183611: contig of 183611 bp in length
            *  183612   183711: gap of unknown length
            *  183712   194613: contig of 10902 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194613
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-344O11"
     misc_feature    1. .1370
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    9001. .9889
                     /note="clone_boundary
                     clone_end:T7
                     site:
                     end_sequence:BZ280306"
     misc_feature    complement(179289. .180103)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:
                     end_sequence:BZ280307"
     misc_feature    181908. .183611
                     /note="wgs_end_extension
                     clone_end:Sp6"
     misc_feature    183712. .185089
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             75.6%;  Score 20.4;  DB 2;  Length 194613;
  Best Local Similarity   95.5%;  Pred. No. 24;
  Matches   21;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          6 CTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||| ||||||
Db      80840 CTAGCTACCCCTAGGCCTAGGC 80819


RESULT 8
AC134931/c
LOCUS       AC134931              124104 bp    DNA     linear   PLN 02-JUN-2004
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 5 clone
            OSJNBb0079L11, complete sequence.
ACCESSION   AC134931
VERSION     AC134931.2  GI:37360982
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 124104)
  AUTHORS   Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Liu,S.-M.,
            Chao,Y.-T., Chang,S.-J., Chen,H.-C., Chen,S.-K., Chen,T.-R.,
            Chen,Y.-L., Cheng,C.-H., Chung,C.-I., Han,S.-Y., Hsiao,S.-H.,
            Hsiung,J.-N., Hsu,C.-H., Huang,J.-J., Kau,P.-I., Lee,M.-C.,
            Leu,H.-L., Li,Y.-F., Lin,S.-J., Lin,Y.-C., Wu,S.-W., Yu,C.-Y.,
            Yu,S.-W., Wu,H.-P. and Shaw,J.-F.
  TITLE     Oryza sativa BAC OSJNBb0079L11 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 124104)
  AUTHORS   Chow,T.-Y. and Hsing,Y.-I.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2002) Institute of Botany, Academia Sinica, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
REFERENCE   3  (bases 1 to 124104)
  AUTHORS   Chow,T.-Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2003) Institute of Botany, Academia Sinica, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
REFERENCE   4  (bases 1 to 124104)
  AUTHORS   Chow,T.-Y. and Hsing,Y.-I.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2004) Institute of Botany, Academia Sinica, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
COMMENT     On Oct 2, 2003 this sequence version replaced gi:23477780.
            The orientation of the sequence is from SP6 to T7 of the BAC clone.
            Genes were predicated from the integrated results of the following:
            BLASTN2.0, BLASTX2.0, GENSCAN (Chris Burge,
            http://genes.mit.edu/GENSCAN.html), Fgenesh
            (http://www.softberry.com/), GlimmerR
            (http://www.tigr.org/softlab/glimmer/glimmer.html), TWINSCAN
            (http://genes.cs.wustl.edu/) and GeneSplicer
            (http://www.tigr.org/tdb/GeneSplicer/index.shtml). The sequence was
            searched against the Swiss-Prot+TrEMBL protein database, the NCBI
            Plant EST database, the TIGR Rice Gene Index and the rice
            full-length cDNA database (KOME,
            http://cdna01.dna.affrc.go.jp/cDNA/). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
```

```
          similarity to other proteins are named after the database hits.
          Genes without significant peptide similarity but with EST
          similarity are named as unknown proteins. Genes without protein or
          EST similarity, that are predicted by more than two gene prediction
          programs over most of their length are annotated as  hypothetical
          proteins. This clone overlaps with P0008A07(accession # AC079021)
          and P0683F12(accession # AC129720).
FEATURES             Location/Qualifiers
     source          1. .124104
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="5"
                     /clone="OSJNBb0079L11"
     gene            complement(2260. .3443)
                     /gene="OSJNBb0079L11.1"
                     /note="unknown protein"
     CDS             complement(join(2260. .2664,3042. .3443))
                     /gene="OSJNBb0079L11.1"
                     /note="'similar to rice EST CA759294, AK105314'"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAT39224.1"
                     /db_xref="GI:47900448"
                     /translation="MAWGLIITPQTFDFIVAHNALRCAKVVLEGKAPKLYEMRANPNC
                     MSNYGYFPLHQAAETFSVDMIKLLLRYGASANLRTSGQKVIEGLLPLHVAIENTCMHK
                     YLEDNLLTDQKHKQVDFDFIYKLVHLLCLPEMKIFLDTTRLLADYTDNVLDELWNYIK
                     EGKLVHAAILLLAAQRKIRTCASSKRNTNHNLSGFDIIKDRITGSIISIETEGPGLTG
                     GKNSKARKQQDDMRMFFRNALVLNSMISEAGEALDEYIQTHSEVQIFLRS"
     gene            12587. .13900
                     /gene="OSJNBb0079L11.2"
                     /note="unknown protein"
     CDS             12587. .13900
                     /gene="OSJNBb0079L11.2"
                     /note="'similar to rice EST D43093, AK063257'"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAT39225.1"
                     /db_xref="GI:47900449"
                     /db_xref="UniProt/TrEMBL:BAC84340"
                     /translation="MCCRRYELKLPIGCDPKQYIVDLILELEGTLEALGDYCFLCQGV
                     WSEFIVGNNLDFLKQLELTSDKLVQVTSNLVRTCYMQKLGTGEEHILKFKYACNGWPT
                     RKFRKMPKNDCNGHWKMSRLFVNLHYFSFGLSTAKTVQKGSCYLRGSDKVLLTKLEMQ
                     SCVNLGDDNVSCHTMAPSEGMDFCELKVLEVTSDGCPIFWILVLAATEWKVGNYTILT
                     EGTYTKMAVCSEQKLQCDALKRNQTGAKHVIQRNTFVAPGNLAYSVTNNYLPEGNKHR
                     TMATRVNVFGHGIHSTENQGFYSEGIQVHSSVQKIPALVLNRELSMDFSKDIAAPMKY
                     ALQVSESFELNRSKKVHDLFHFDDGLLSSSQELEHPWDPGGLLHRLGDKTIFKERGLL
                     GPDGLWLDRYNNPPIQPKKAQLERQQKQQQQEGWGHRLITWTWWS"
     gene            28453. .38461
                     /gene="OSJNBb0079L11.3"
                     /note="putative DNA-binding protein"
     CDS             join(28453. .28479,29198. .29278,29520. .29562,29766. .29844,
                     29953. .30112,30227. .30354,30796. .30864,31458. .31545,
                     32178. .32253,32429. .32502,34297. .34404,35445. .35606,
                     35841. .35945,36109. .36189,36265. .37465,38316. .38461)
                     /gene="OSJNBb0079L11.3"
                     /note="'similar to rice EST CB685703, AK105290 '"
                     /codon_start=1
                     /product="putative DNA-binding protein"
                     /protein_id="AAT39226.1"
                     /db_xref="GI:47900450"
                     /db_xref="UniProt/TrEMBL:Q9FKK4"
                     /translation="MADLVSSCKDKLAYFRIKELKDILNQLGLPKQGKKQDLIDRVLA
                     LLTDEQGQRHHGWGRKNSLTKEAVAKIVDDTYRKMQIQCAPDLATRSHSGSDFSFRPI
                     EEAYDSFQPEAKVRCICSSTMVNDSMIQCEDQRCQVWQHLNCVLIPDKPGESAEVPPV
                     FYCELCRLSRADPFWVTAGNPLLPVKFVSSGVTNDGTSVPQSVEKSFQLSRSDRETVQ
                     RQEYDLQVWCMLLNDKVQFRMQWPQYAELHVNGISVRVVTRPGSQLLGINGRDDGPLI
                     TTCSREGINKICLSRVDARTFCFGVRIAKRRTVAQVLNLVPKEAEGESFEHALARVRR
                     CLGGGDTAENADSDSDLEVVAESVTVNLRCPNSGSRMRIAGRFKPCIHMGCFDLETFV
                     ELNQRSRKWQCPICLKNYSLESLMIDPYFNRITSLLRNCNEDVNEVDVKPDGSWRVKG
                     DAASRELSQWHMPDGTLCNPKEDVKPAMQNGNEQMMEGTSDGQKSLKIGIKRNPNGIW
                     EVSSKADDKKPSVVGNRMQNNSGFRALNNIMHMSNSPTSSYRDGEDPSVNQESNRHVD
                     LSLNNGNNEFDSFSLNFGQACNTDDRPQQQHNATDVIVLSDSDEENDAMVCPPAVYDN
                     TTTANGSGFPFTTNGIGYTERYQEDAGVGTSGLGLLSNNVDDFEMNNWQMHSSYQQPE
                     QGFQFFGNDTDVHNTFVGSHNSFGLAPNDYSLDCNVGVEEASVTPALSVCRNSNEMHG
                     SLVDNPLALVGDDPSLQIFLPSQPSSVPLQEELSERANAPNGVQSDDWISLTLAAGGG
                     GNEEPAPADVNSQPQIPSTETGIEPLTDAASAFLSTNIERRSGADLNPRRIENIFSHP
                     RQPRSVRPRLCLSIDTDSE"
     gene            complement(42534. .43691)
                     /gene="OSJNBb0079L11.4"
                     /note="hypothetical protein"
     CDS             complement(42534. .43691)
                     /gene="OSJNBb0079L11.4"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAT39227.1"
                     /db_xref="GI:47900451"
                     /translation="MKPPAASPGRAEKQLPAPPGLARLLLSKSRRGGGGGRSRRAPAT
                     SPMFVSRGRGGAGRAAAAAADGEPSSPKVTCIGQVRMRKGKKGKKAAAASKVAAPEKG
                     RGYCRCLKKAFLCGGLFDFDRRRRPKAPSPEVVERSRRSPWVFSSRDVAVAAAPKQPD
                     PRGEHAEEDEEMEASAGGVFRSLEGEEAERLGINGGGDDKEEDEEEEEEEAELVSSAT
                     TTPPKNALLLMRCRSAPQNRSSPLTSRFPVAAAPSPTKDAAAAAEIAAPSPRKSQKAS
                     AAAEAQHEERQEEMASVQEQEEVREEEEEDVDDDDDYDEEDMRCSSARPLVLPRCKSE
                     PATTAAAKMAGGTAADATTAGCFWANGGSSGRRRHAPAAAAAATPVALAGH"
     gene            55992. .58307
                     /gene="OSJNBb0079L11.5"
                     /note="putative receptor like protein kinase LRK1"
     CDS             55992. .58307
                     /gene="OSJNBb0079L11.5"
                     /note="similar to rice EST AK120637"
                     /codon_start=1
                     /product="putative receptor like protein kinase LRK1"
                     /protein_id="AAT39228.1"
                     /db_xref="GI:47900452"
                     /db_xref="UniProt/TrEMBL:Q9LYX1"
                     /translation="MPPPMAPPPPPPPPPPILLLLLLLLLAAVAADVSTALCFDYATLT
                     LGSLKLLGDAHLKNGTIRLSRDLPVPNSGAGRALYATPVALRGGFSTQFAFTVATLNA
                     DSVGGGLAFVLASDGVTLGDAGPYIGVSAASDVAAVEFDTLMDVQFGDVNGNHVGLDL
                     GSMVSAAVADLDGVGVELTSGRTVNAWIEYSPKSGMEVFVSYSPKRPAEPVLSAPLDL
                     GEYVKGDAFVGFSASTQGSTEMHAVEWWTFSTPTSSSSSPSKPSPRMATPSSPPPEAP
                     VSSAAPPPASLNPMLPSPPQLPGVSTTTPSPPASTVSMPPTNSVAVASAPANSTAGIS
                     NAGSPHPPAHAAVAGAATAGAFVAVSFAGFALWALARRARARRRGTTALAAVADKRDS
                     LASAAALARSPREFTYKELSAATRGFDASRVIGNGAFGTVYKGIIPDTGAMVAVKRCT
                     NASADGAQARSEFLSELSIIAGLRHRNLLRLQGWCHEKGEILLVYDYMRNGSLDKALF
                     DASSPVLPWSHRREILAGVASALAYLHHECERRVIHRDVKSSNVMLDDAYRARLGDFG
                     LARQAEHGESPDATAAAGTMGYLAPEYLLTGRATEATDVFSFGALVLEVACGRRPIGA
                     TEGRCNNLVEWVWSLHGAGQVLDAVDARLRGEYDEAEMRRAMLVGLACSSPEALRPG
                     MRAVVQMLGGEADPPFVPAARPSMSFSANHQLLLSLQDSVSDYNALGLNDLSDDSSSD
                     SLSSSSLTSTLRKGGHDIAAFSSAAAGDAAR"
     gene            62625. .65043
                     /gene="OSJNBb0079L11.6"
                     /note="hypothetical protein"
     CDS             join(62625. .62745,63333. .63550,63645. .63834,64053. .64444,
                     64633. .65043)
                     /gene="OSJNBb0079L11.6"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAT39229.1"
                     /db_xref="GI:47900453"
                     /db_xref="UniProt/TrEMBL:Q9SFT7"
                     /translation="MSFFSCFKPDKKMLSKRMEEMPFTVVKKASSQHGSSLKNSESDK
                     SPRGHSNNKKSSVAAKNTEPPKRIPITAKAERSFTFRELATATNNFHPDCIVGEGGFG
                     RVYKGQLEDGQVVAVKQMERNGFQGNREFLIEVMILGHLNHPNLVNLVGYCSDGDQRL
                     LAYEYMALGSLADHLLDITPDQEPLSWRTRMKIAHGTAKGLEHLHEKMSPPVIYRDLK
                     SPNILLDKDYNPKLSDFGLAKLGPFEGDKHVSTRVMGTFGYCAPEYVRTGMLSTKTDV
                     YSFGVFLLELITGRRAVDTCRPVCEQILAYWAKPMLHDRRRYHELVDPLLRGDYPDKD
                     FNQAAAVAAICIEDEASVRPYMSDIVVALGFLAEVPAGCEERINAEPQNRKDEDPSVT
                     GNTKKDQSTFDRQRAVAEAIEWGSVRQKQMAQIKEKKAQPQGIVAPKTNKL"
     gene            68067. .70231
                     /gene="OSJNBb0079L11.7"
                     /note="hypothetical protein"
     CDS             join(68067. .68175,68472. .68689,68790. .68979,69242. .69633,
                     69818. .70231)
                     /gene="OSJNBb0079L11.7"
                     /codon_start=1
```

```
                     /product="hypothetical protein"
                     /protein_id="AAT39230.1"
                     /db_xref="GI:47900454"
                     /db_xref="UniProt/TrEMBL:Q9SFT7"
                     /translation="MSCFSCFKPEKKMKSKGMEAREVTVVKKHGASLKNSESEKLPCV

  Query Match             74.8%;  Score 20.2;  DB 8;  Length 124104;
  Best Local Similarity   88.0%;  Pred. No. 32;
  Matches   22;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AGCCTAGCTACCCCTAGGTCTAGGC 27
              | ||||||||||||| || ||||||
Db       6726 ACCCTAGCTACCCCTGGGGCTAGGC 6702


RESULT 9
AC134515
LOCUS       AC134515               119743 bp    DNA     linear   HTG 27-SEP-2002
DEFINITION  Rattus norvegicus clone CH230-10I6, *** SEQUENCING IN PROGRESS ***,
            63 unordered pieces.
ACCESSION   AC134515
VERSION     AC134515.1  GI:23334689
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 119743)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119743)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ---------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
              Center project name: GDSS
              Center clone name: CH230-10I6
            ---------- Summary Statistics
              Sequencing vector: Plasmid;
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 65609 bases at least Q40
              Consensus quality: 70531 bases at least Q30
              Consensus quality: 73885 bases at least Q20
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 63 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1021: contig of 1021 bp in length
            *     1022     1121: gap of unknown length
            *     1122     2410: contig of 1289 bp in length
            *     2411     2510: gap of unknown length
            *     2511     3615: contig of 1105 bp in length
            *     3616     3715: gap of unknown length
            *     3716     5208: contig of 1493 bp in length
            *     5209     5308: gap of unknown length
            *     5309     6348: contig of 1040 bp in length
            *     6349     6448: gap of unknown length
            *     6449     7556: contig of 1108 bp in length
            *     7557     7656: gap of unknown length
            *     7657     8690: contig of 1034 bp in length
            *     8691     8790: gap of unknown length
            *     8791     9982: contig of 1192 bp in length
            *     9983    10082: gap of unknown length
            *    10083    11108: contig of 1026 bp in length
            *    11109    11208: gap of unknown length
            *    11209    12381: contig of 1173 bp in length
            *    12382    12481: gap of unknown length
            *    12482    13944: contig of 1463 bp in length
            *    13945    14044: gap of unknown length
            *    14045    15293: contig of 1249 bp in length
            *    15294    15393: gap of unknown length
            *    15394    16836: contig of 1443 bp in length
            *    16837    16936: gap of unknown length
            *    16937    18267: contig of 1331 bp in length
            *    18268    18367: gap of unknown length
            *    18368    19813: contig of 1446 bp in length
            *    19814    19913: gap of unknown length
            *    19914    21224: contig of 1311 bp in length
            *    21225    21324: gap of unknown length
            *    21325    22712: contig of 1388 bp in length
            *    22713    22812: gap of unknown length
            *    22813    23831: contig of 1019 bp in length
```

```
  *    23832    23931: gap of unknown length
  *    23932    25886: contig of 1955 bp in length
  *    25887    25986: gap of unknown length
  *    25987    27306: contig of 1320 bp in length
  *    27307    27406: gap of unknown length
  *    27407    28973: contig of 1567 bp in length
  *    28974    29073: gap of unknown length
  *    29074    30417: contig of 1344 bp in length
  *    30418    30517: gap of unknown length
  *    30518    31979: contig of 1462 bp in length
  *    31980    32079: gap of unknown length
  *    32080    33446: contig of 1367 bp in length
  *    33447    33546: gap of unknown length
  *    33547    35631: contig of 2085 bp in length
  *    35632    35731: gap of unknown length
  *    35732    37564: contig of 1833 bp in length
  *    37565    37664: gap of unknown length
  *    37665    38768: contig of 1104 bp in length
  *    38769    38868: gap of unknown length
  *    38869    39951: contig of 1083 bp in length
  *    39952    40051: gap of unknown length
  *    40052    41348: contig of 1297 bp in length
  *    41349    41448: gap of unknown length
  *    41449    42876: contig of 1428 bp in length
  *    42877    42976: gap of unknown length
  *    42977    44226: contig of 1250 bp in length
  *    44227    44326: gap of unknown length
  *    44327    45971: contig of 1645 bp in length
  *    45972    46071: gap of unknown length
  *    46072    48011: contig of 1940 bp in length
  *    48012    48111: gap of unknown length
  *    48112    49629: contig of 1518 bp in length
  *    49630    49729: gap of unknown length
  *    49730    51576: contig of 1847 bp in length
  *    51577    51676: gap of unknown length
  *    51677    53467: contig of 1791 bp in length
  *    53468    53567: gap of unknown length
  *    53568    54758: contig of 1191 bp in length
  *    54759    54858: gap of unknown length
  *    54859    56642: contig of 1784 bp in length
  *    56643    56742: gap of unknown length
  *    56743    58752: contig of 2010 bp in length
  *    58753    58852: gap of unknown length
  *    58853    60328: contig of 1476 bp in length
  *    60329    60428: gap of unknown length
  *    60429    62184: contig of 1756 bp in length
  *    62185    62284: gap of unknown length
  *    62285    64491: contig of 2207 bp in length
  *    64492    64591: gap of unknown length
  *    64592    66173: contig of 1582 bp in length
  *    66174    66273: gap of unknown length
  *    66274    67696: contig of 1423 bp in length
  *    67697    67796: gap of unknown length
  *    67797    69841: contig of 2045 bp in length
  *    69842    69941: gap of unknown length
  *    69942    71536: contig of 1595 bp in length
  *    71537    71636: gap of unknown length
  *    71637    73791: contig of 2155 bp in length
  *    73792    73891: gap of unknown length
  *    73892    76129: contig of 2238 bp in length
  *    76130    76229: gap of unknown length
  *    76230    78594: contig of 2365 bp in length
  *    78595    78694: gap of unknown length
  *    78695    80868: contig of 2174 bp in length
  *    80869    80968: gap of unknown length
  *    80969    82506: contig of 1538 bp in length
  *    82507    82606: gap of unknown length
  *    82607    85214: contig of 2608 bp in length
  *    85215    85314: gap of unknown length
  *    85315    87565: contig of 2251 bp in length
  *    87566    87665: gap of unknown length

Query Match             73.3%;  Score 19.8;  DB 2;  Length 119743;
  Best Local Similarity   91.3%;  Pred. No. 50;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CCTAGCTACCCCTAGGTCTAGGC 27
              |||| |||||||||| ||||||
Db      93195 CCTAACTACCCCTAGGACTAGGC 93217


RESULT 10
AC103070
LOCUS       AC103070              211465 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-171N16, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC103070
VERSION     AC103070.6  GI:30579789
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 211465)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 211465)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 211465)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:23321751.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJDL
                Center clone name: CH230-171N16
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 165184 bases at least Q40
                Consensus quality: 169724 bases at least Q30
                Consensus quality: 173497 bases at least Q20
                Estimated insert size: 175743; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    15062: contig of 15062 bp in length
            *   15063    15162: gap of unknown length
            *   15163   211465: contig of 196303 bp in length.
FEATURES             Location/Qualifiers
     source          1. .211465
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-171N16"
     misc_feature    9750. .11143
                     /note="wgs_contig"
     misc_feature    31208. .33466
                     /note="wgs_contig"
     misc_feature    136133. .137193
                     /note="wgs_contig"
     misc_feature    143435. .145324
                     /note="wgs_contig"
     misc_feature    184822. .187065
                     /note="wgs_contig"
     misc_feature    189375. .189980
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH289014"
     misc_feature    209353. .211465
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             73.3%;  Score 19.8;  DB 2;  Length 211465;
  Best Local Similarity   91.3%;  Pred. No. 48;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CCTAGCTACCCCTAGGTCTAGGC 27
              |||| |||||||||||  ||||||
Db      22022 CCTAACTACCCCTAGGACTAGGC 22044


RESULT 11
AC095896
LOCUS       AC095896              212559 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-10N20, *** SEQUENCING IN PROGRESS
            ***, 10 unordered pieces.
ACCESSION   AC095896
VERSION     AC095896.6  GI:24940801
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 212559)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egán,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
```

```
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212559)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 212559)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:23269609.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDUL
                Center clone name: CH230-10N20
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 172337 bases at least Q40
                Consensus quality: 177428 bases at least Q30
                Consensus quality: 180480 bases at least Q20
                Estimated insert size: 178337; sum-of-contigs estimation
                Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    45798: contig of 45798 bp in length
            *    45799    45898: gap of unknown length
            *    45899   146268: contig of 100370 bp in length
            *   146269   146368: gap of unknown length
            *   146369   150042: contig of 3674 bp in length
            *   150043   150142: gap of unknown length
            *   150143   162331: contig of 12189 bp in length
            *   162332   162431: gap of unknown length
            *   162432   199327: contig of 36896 bp in length
            *   199328   199427: gap of unknown length
            *   199428   200818: contig of 1391 bp in length
            *   200819   200918: gap of unknown length
            *   200919   202031: contig of 1113 bp in length
            *   202032   202131: gap of unknown length
            *   202132   203608: contig of 1477 bp in length
            *   203609   203708: gap of unknown length
            *   203709   206315: contig of 2607 bp in length
            *   206316   206415: gap of unknown length
            *   206416   212559: contig of 6144 bp in length.
FEATURES             Location/Qualifiers
     source          1. .212559
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-10N20"
     misc_feature    129775. .131842
                     /note="wgs_contig"
     misc_feature    146369. .147484
                     /note="wgs_contig"
     misc_feature    162432. .164020
                     /note="wgs_contig"
ORIGIN

  Query Match             73.3%;  Score 19.8;  DB 2;  Length 212559;
  Best Local Similarity   91.3%;  Pred. No. 48;
  Matches   21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          5 CCTAGCTACCCCTAGGTCTAGGC 27
              |||| ||| ||||||||||||||
Db      90486 CCTAACTATCCCTAGGTCTAGGC 90508


RESULT 12
SSAPOA11
LOCUS       SSAPOA11                 857 bp    DNA     linear   MAM 18-APR-1994
DEFINITION  S.scrofa gene for apolipoprotein A1, exons 1-3.
ACCESSION   X69478
VERSION     X69478.1  GI:1887
KEYWORDS    apolipoprotein A-I; lecithin cholesterol acyltransferase cofactor;
            lipid binding.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 857)
  AUTHORS   Mockel,B., Zinke,H., Flach,R., Weiss,B., Weiler-Guttler,H. and
            Gassen,H.G.
  TITLE     Expression of apolipoprotein A-I in porcine brain endothelium in
            vitro
  JOURNAL   J. Neurochem. 62 (2), 788-798 (1994)
  MEDLINE   94125128
   PUBMED   8294940
REFERENCE   2  (bases 1 to 857)
  AUTHORS   Moeckel,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-1992) B. Moeckel, Inst. fuer Biochemie, Prof.
            Gassen, Technische Hochshule Darmstadt, Petersenstr. 22, 6100
            Darmstadt, FRG
FEATURES             Location/Qualifiers
     source          1. .857
                     /organism="Sus scrofa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9823"
                     /clone="PCR products"
                     /tissue_type="liver"
                     /dev_stage="adult"
     CAAT_signal     70. .75
     TATA_signal     151. .156
     gene            join(179. .857,X69479.1:424. .1068)
                     /gene="APOA1"
     mRNA            join(179. .199,369. .431,612. .765,X69479.1:424. .>1068)
                     /gene="APOA1"
                     /label=apo_mrna
     exon            179. .199
                     /gene="APOA1"
                     /number=1
     intron          200. .368
```

```
                     /gene="APOA1"
                     /number=1
     exon            369. .431
                     /gene="APOA1"
                     /number=2
     intron          432. .611
                     /gene="APOA1"
                     /number=2
     exon            612. .765
                     /gene="APOA1"
                     /number=3
     intron          766. .>857
                     /gene="APOA1"
                     /number=3
ORIGIN

  Query Match             72.6%;  Score 19.6;  DB 4;  Length 857;
  Best Local Similarity   84.6%;  Pred. No. 90;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
              || ||||||| |||| || |||||||
Db        224 TACCCTAGCTTCCCCCAGCTCTAGGC 249


RESULT 13
AK173314
LOCUS       AK173314               5551 bp    mRNA    linear   ROD 28-JUL-2004
DEFINITION  Mus musculus mRNA for mKIAA1973 protein.
ACCESSION   AK173314
VERSION     AK173314.1  GI:50511212
KEYWORDS    FLI_CDNA.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Okazaki,N., Kikuno,R.F., Ohara,R., Inamoto,S., Koseki,H.,
            Hiraoka,S., Saga,Y., Seino,S., Nishimura,M., Kaisho,T., Hoshino,K.,
            Kitamura,H., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Prediction of the Coding Sequences of Mouse Homologues of KIAA
            Gene: IV. The Complete Nucleotide Sequences of 500 Mouse
            KIAA-Homologous cDNAs Identified by Screening of Terminal Sequences
            of cDNA Clones Randomly Sampled from Size-Fractionated Libraries
  JOURNAL   DNA Res. 11, 205-218 (2004)
REFERENCE   2  (bases 1 to 5551)
  AUTHORS   Okazaki,N., Kikuno,R.F., Nagase,T., Ohara,O. and Koga,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAY-2004) Hisashi Koga, Kazusa DNA Research
            Institute, Laboratory for Genome Informatics; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:mouse@kazusa.or.jp, Tel:81-438-52-3919, Fax:81-438-52-3918)
COMMENT     The CREATE program supported by Japan science and technology
            corporation; cDNA full insert sequencing: Kazusa DNA Research
            Institute; cDNA library construction, clone selection and 5'- &
            3'-end one pass sequencing.
FEATURES             Location/Qualifiers
     source          1. .5551
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="mfj04206"
                     /note="vector:modified pBC SK+"
     gene            1. .5551
                     /gene="mKIAA1973"
     CDS             <1. .3097
                     /gene="mKIAA1973"
                     /note="CDS is predicted by in silico analysis. Start codon
                     is not identified."
                     /codon_start=2
                     /evidence=not_experimental
                     /product="mKIAA1973 protein"
                     /protein_id="BAD32592.1"
                     /db_xref="GI:50511213"
                     /translation="RGPPGSRKLGEGAGTETLWPRDALLFAVENLNRVEGLLPYNLSL
                     EVVMAIEAGLGDLPLMPFSSPSSPWSSDPFSFLQSVCHTVVVQGVSALLAFPQSQGEM
                     MELDLVSSVLHIPVLSIVRHEFPRESQNPLHLQLSLENSLSSDADVTVSILTMNNWYN
                     FSLLLCQEDWNITDFLLLTENNSKFHLESIINITANLSSTKDLLSFLQVQLENIRNST
                     PTMVMFGCDMGSIRQIFEMSTQFGLSPPDLHWVLGDSQNVEELRTEGLPLGLIAHGKT
                     TQSVFEYYVQDAMELVARAVATATMIQPELALLPSTMNCMDVKTTNLTSGQYLSRFLA
                     NTTFRGLSGSIKVKGSTIVSSENNFFIWNLQYDPMGKPMWTRLGSWQGGRIVMDSGIW
                     PEQAQRHKTHFHHPNKLHLRVVTLIEHPFVFTREVDDEGLCPAGQLCLDPMTNDSSIL
                     DSLFSSLHSSNDTVPIKFKKCCYGYCIDLLEQLAEDMNFDFDLYIVGDGKYGAWKNGH
                     WTGLVGDLLSGTANMAVTSFSINTARSQVIDFTSPFFSTSLGILVRTRDTAAPIGAFM
                     WPLHWTMWLGIFVALHITAIFLTLYEWKSPFGMTPKGRNRNKVFSFSSALNVCYALLF
                     GRTAAIKPPKCWTGRFLMNLWAIFCMFCLSTYTANLAAVMVGEKIYEELSGIHDPKLH
                     HPSQGFRFGTVRESSAEDYVRQSFPEMHEYMRRYNVPATPDGVQYLKNDPEKLDAFIM
                     DKALLDYEVSIDADCKLLTVGKPFAIEGYGIGLPPNSPLTSNISELISQYKSHGFMDV
                     LHDKWYKVVPCGKRSFAVTETLQMGIKHFSGLFVLLCIGFGLSILTTIGEHIVYRLLL
                     PRIKNKSKLQYWLHTSQRFHRALNTSFVEEKQPCSKTKRVEKSRWRRWTCKTEGDSEL
                     SLFPRSNMGPQQLMVWNTSNLSHDNQRKYIFNDEEGQNQLGTQTHQDIPLPPRRRELP
                     ASLTTNGKADSLNVARNSVMQELSELEKQIQVIRQELQLAVSRKTELEEYQRTNRTCE
                     S"
ORIGIN

  Query Match             72.6%;  Score 19.6;  DB 10;  Length 5551;
  Best Local Similarity   84.6%;  Pred. No. 79;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||| |||||  || |||||
Db       4186 TAGCCTAGCTTCCCCTTTGTTTAGGC 4211


RESULT 14
AL732521/c
LOCUS       AL732521               207223 bp    DNA     linear   ROD 16-JUL-2003
DEFINITION  Mouse DNA sequence from clone RP23-134A17 on chromosome 4, complete
            sequence.
ACCESSION   AL732521
VERSION     AL732521.18  GI:32879627
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 207223)
  AUTHORS   Wallis,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 16, 2003 this sequence version replaced gi:32480449.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
```

occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-134A17 is from the RPCI-23 Mouse BAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

```
FEATURES             Location/Qualifiers
     source          1. .207223
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="4"
                     /clone="RP23-134A17"
                     /clone_lib="RPCI-23"
ORIGIN

  Query Match             72.6%;  Score 19.6;  DB 10;  Length 207223;
  Best Local Similarity   84.6%;  Pred. No. 61;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
              ||||||||||| |||||  || |||||
Db     117297 TAGCCTAGCTTCCCCTTTGTTTAGGC 117272


RESULT 15
AC126959
LOCUS       AC126959               212481 bp    DNA     linear   ROD 12-AUG-2004
DEFINITION  Rattus norvegicus 8 BAC CH230-10P12 (Children's Hospital Oakland
            Research Institute) complete sequence.
ACCESSION   AC126959
VERSION     AC126959.6  GI:49170123
KEYWORDS    HTG.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 212481)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
            Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
            Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
            Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
            Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
            Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
            Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
            Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
            Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
            Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
            Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
            Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
            Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A.,
            Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
            Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
            Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
            Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
            Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
            Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
            Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
            Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 212481)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 212481)
  AUTHORS   .
  CONSRTM   Rat Genome Sequencing Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 212481)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-2004) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 24, 2004 this sequence version replaced gi:30522732.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.
FEATURES             Location/Qualifiers
     source          1. .212481
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /chromosome="8"
                     /clone="CH230-10P12"
     repeat_region   complement(177. .364)
                     /rpt_family="MIRb"
     repeat_region   complement(730. .937)
                     /rpt_family="B3"
     repeat_region   complement(1873. .2096)
                     /rpt_family="MLT1D"
     repeat_region   3222. .3277
                     /rpt_family="(CAA)n"
     repeat_region   3877. .4035
                     /rpt_family="MIRb"
     repeat_region   complement(4624. .4772)
                     /rpt_family="MLT2B"
     repeat_region   complement(4953. .5052)
                     /rpt_family="URR1B"
     repeat_region   5844. .5896
                     /rpt_family="(TATAA)n"
     repeat_region   6289. .6367
                     /rpt_family="polypyrimidine"
     repeat_region   complement(6435. .6758)
                     /rpt_family="L1MEc"
     repeat_region   complement(6759. .6966)
```

```
                     /rpt_family="B3A"
  repeat_region      complement(6967. .7615)
                     /rpt_family="L1MEc"
  repeat_region      7705. .7765
                     /rpt_family="(TCCC)n"
  repeat_region      8202. .8242
                     /rpt_family="(TTA)n"
  repeat_region      complement(9048. .9143)
                     /rpt_family="ID_Rn"
  repeat_region      10229. .10329
                     /rpt_family="PB1D10"
  repeat_region      10466. .10487
                     /rpt_family="AT_rich"
  repeat_region      11686. .11732
                     /rpt_family="(TTTA)n"
  repeat_region      12682. .12759
                     /rpt_family="(TA)n"
  repeat_region      14410. .14920
                     /rpt_family="MLT2B2"
  repeat_region      15871. .16042
                     /rpt_family="MIRb"
  repeat_region      16158. .16262
                     /rpt_family="MER3"
  repeat_region      17092. .17153
                     /rpt_family="(TC)n"
  repeat_region      17267. .17428
                     /rpt_family="CT-rich"
  repeat_region      complement(17881. .18094)
                     /rpt_family="MIRb"
  repeat_region      20587. .20660
                     /rpt_family="MER46C"
  repeat_region      complement(20862. .20992)
                     /rpt_family="MIRb"
  repeat_region      21236. .21275
                     /rpt_family="(TA)n"
  repeat_region      complement(22115. .22273)
                     /rpt_family="L2"
  repeat_region      complement(24710. .24811)
                     /rpt_family="MLT1I"
  repeat_region      complement(25815. .25888)
                     /rpt_family="MIRb"
  repeat_region      complement(26360. .26455)
                     /rpt_family="ID_Rn"
  repeat_region      27537. .27653
                     /rpt_family="(CA)n"
  repeat_region      complement(28087. .28146)
                     /rpt_family="MLT1B"
  repeat_region      28152. .28515
                     /rpt_family="MTD"
  repeat_region      complement(28560. .28740)
                     /rpt_family="MLT1A0"
  repeat_region      complement(29743. .30243)
                     /rpt_family="RMER15"
  repeat_region      complement(30256. .30474)
                     /rpt_family="URR1B"
  repeat_region      31301. .31370
                     /rpt_family="(TATAG)n"
  repeat_region      complement(31442. .31500)
                     /rpt_family="RMER21A"
  repeat_region      complement(31639. .31926)
                     /rpt_family="RMER21B"
  repeat_region      complement(31964. .32097)
                     /rpt_family="MTD"
  repeat_region      complement(32153. .32410)
                     /rpt_family="MTD"
  repeat_region      33398. .33753
                     /rpt_family="MTD"
  repeat_region      complement(33851. .33933)
                     /rpt_family="L2"
  repeat_region      34224. .34417
                     /rpt_family="B3"
  repeat_region      34420. .34487
                     /rpt_family="(CA)n"
  repeat_region      34488. .34542
                     /rpt_family="(GA)n"
  repeat_region      35746. .35799
                     /rpt_family="(TC)n"
  repeat_region      35803. .35857
                     /rpt_family="(CA)n"
  repeat_region      36195. .36239
                     /rpt_family="(TA)n"
  repeat_region      complement(36740. .37368)
                     /rpt_family="L1MA9"
  repeat_region      40383. .40406
                     /rpt_family="(TA)n"

  Query Match             72.6%;  Score 19.6;  DB 10;  Length 212481;
  Best Local Similarity   84.6%;  Pred. No. 61;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
              ||  ||| ||| ||||||||||||||
Db      91228 TATTCTATCTATCCCTAGGTCTAGGC 91253


Search completed: October 13, 2005, 19:37:32
Job time : 1488 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 13, 2005, 17:44:53 ; Search time 1122 Seconds
(without alignments)
142.454 Million cell updates/sec

Title: US-09-744-097A-76
Perfect score: 27
Sequence: 1 gtagcctagctacccctaggtctaggc 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		27	100.0	27	3	AAA32245	Aaa32245 Green ter
c 2		27	100.0	27	3	AAA32253	Aaa32253 Distal sp
c 3		27	100.0	27	3	AAA32261	Aaa32261 Double di
c 4		27	100.0	27	3	AAA32262	Aaa32262 Double di
c 5		27	100.0	27	3	AAA32274	Aaa32274 One to tw
c 6		27	100.0	27	3	AAA32293	Aaa32293 Green set
7		27	100.0	27	3	AAA32255	Aaa32255 Proximal
c 8		27	100.0	27	3	AAA32243	Aaa32243 Distal sp
9		27	100.0	27	3	AAA32264	Aaa32264 Proximal
10		27	100.0	27	3	AAA32241	Aaa32241 Proximal
c 11		27	100.0	27	3	AAA32275	Aaa32275 One to tw
c 12		27	100.0	27	4	AAS14192	Aas14192 GeneTAG F
c 13		27	100.0	27	4	AAS14194	Aas14194 First-GRE
14		27	100.0	27	4	AAS14183	Aas14183 Fragment
15		27	100.0	27	4	AAS14198	Aas14198 First-GRE
16		27	100.0	27	4	AAS14191	Aas14191 First-GRE
c 17		27	100.0	27	4	AAS14181	Aas14181 Fragment
c 18		27	100.0	29	4	AAS14197	Aas14197 First-GRE
c 19		27	100.0	31	4	AAS14190	Aas14190 First-GRE
c 20		27	100.0	36	4	AAS14195	Aas14195 First-GRE
c 21		27	100.0	42	3	AAA32311	Aaa32311 Anti-sens
c 22		27	100.0	42	4	AAS14178	Aas14178 Modified
c 23		27	100.0	57	3	AAA32233	Aaa32233 Green rep
c 24		27	100.0	72	3	AAA32310	Aaa32310 Anti-sens
c 25		25	92.6	25	3	AAA32225	Aaa32225 Terminato
c 26		25	92.6	25	3	AAA32188	Aaa32188 Terminato
27		25	92.6	25	3	AAA32177	Aaa32177 Distal li
28		25	92.6	37	3	AAA32184	Aaa32184 Overlap l
29		25	92.6	37	3	AAA32174	Aaa32174 Overlap o
30		25	92.6	49	3	AAA32220	Aaa32220 Reporter
c 31		25	92.6	57	3	AAA32222	Aaa32222 Proximal
32		25	92.6	57	3	AAA32223	Aaa32223 Distal su
33		25	92.6	57	3	AAA32186	Aaa32186 Distal su
c 34		25	92.6	57	3	AAA32185	Aaa32185 Proximal
35		25	92.6	63	3	AAA32173	Aaa32173 Target sp
36		25	92.6	63	3	AAA32183	Aaa32183 Target ol
c 37		24	88.9	24	3	AAA32175	Aaa32175 Proximal
c 38		24	88.9	24	3	AAA32182	Aaa32182 Terminato
c 39		24	88.9	24	3	AAA32179	Aaa32179 Proximal
c 40		24	88.9	40	3	AAA32199	Aaa32199 Second re
c 41		24	88.9	40	3	AAA32191	Aaa32191 Forward r
c 42		24	88.9	40	3	AAA32193	Aaa32193 Reverse r
c 43		24	88.9	40	3	AAA32204	Aaa32204 Forward r
c 44		24	88.9	40	3	AAA32197	Aaa32197 Reporter
c 45		24	88.9	40	3	AAA32206	Aaa32206 Reverse r

ALIGNMENTS

RESULT 1
AAA32245
ID AAA32245 standard; DNA; 27 BP.
XX
AC AAA32245;
XX
DT 14-JUL-2000 (first entry)
XX
DE Green terminator sequence used in gene-tag reporter construction.
XX
KW Gene-tag reporter; detection; gene mapping; mutation identification;
KW cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS Synthetic.
XX
PN WO200004192-A1.
XX
PD 27-JAN-2000.
XX
PF 16-JUL-1999; 99WO-US016242.
XX
PR 17-JUL-1998; 98US-0093219P.
XX
PA (UYEM-) UNIV EMORY.
XX
PI Shafer DA;
XX
DR WPI; 2000-182448/16.
XX
PT New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT mapping genes or mutational analysis.
XX
PS Example 12; Page 62; 164pp; English.
XX
CC This sequence is used in the construction of the gene-tag reporter of the
CC invention. The invention relates to a gene-tag reporter for joining a
CC probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC comprises a labelled double-stranded polynucleotide sequence having one
CC or more linkers that comprise a single stranded sequence hybridisable to
CC a complement but not to the target probe. Also included in the invention
CC is a reporter array, comprising at least two gene-tag reporters linked
CC end to end by hybridisable linkers. The reporters are used in various new
CC methods for detecting and mapping genes; identifying mutations and

```
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 5 A; 9 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27


RESULT 2
AAA32253/c
ID   AAA32253 standard; DNA; 27 BP.
XX
AC   AAA32253;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Distal spacer oligomer A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 13; Page 67; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 3
AAA32261/c
ID   AAA32261 standard; DNA; 27 BP.
XX
AC   AAA32261;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Double distal spacer oligomer A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 14; Page 69; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
```

```
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 4
AAA32262/c
ID   AAA32262 standard; DNA; 27 BP.
XX
AC   AAA32262;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Double distal spacer oligomer B used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 14; Page 69; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 5
AAA32274/c
ID   AAA32274 standard; DNA; 27 BP.
XX
AC   AAA32274;
XX
DT   14-JUL-2000  (first entry)
XX
DE   One to two multilinker A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 15; Page 70; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 6
AAA32293/c
ID   AAA32293 standard; DNA; 27 BP.
XX
AC   AAA32293;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Green set first linker A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
```

```
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 16; Page 72; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 7
AAA32255
ID   AAA32255 standard; DNA; 27 BP.
XX
AC   AAA32255;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Proximal spacer oligomer A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 13; Page 67; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 5 A; 9 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27


RESULT 8
AAA32243/c
ID   AAA32243 standard; DNA; 27 BP.
XX
AC   AAA32243;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Distal spacer oligomer A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 12; Page 62; 164pp; English.
```

```
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match          100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity  100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 9
AAA32264
ID   AAA32264 standard; DNA; 27 BP.
XX
AC   AAA32264;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Proximal spacer oligomer A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 14; Page 69; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 5 A; 9 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match          100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity  100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27


RESULT 10
AAA32241
ID   AAA32241 standard; DNA; 27 BP.
XX
AC   AAA32241;
XX
DT   14-JUL-2000  (first entry)
XX
DE   Proximal spacer oligomer A used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 12; Page 62; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
```

```
CC   amplification
XX
SQ   Sequence 27 BP; 5 A; 9 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity  100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27


RESULT 11
AAA32275/c
ID   AAA32275 standard; DNA; 27 BP.
XX
AC   AAA32275;
XX
DT   14-JUL-2000  (first entry)
XX
DE   One to two multilinker B used in gene-tag reporter construction.
XX
KW   Gene-tag reporter; detection; gene mapping; mutation identification;
KW   cancer; mutant virus; human diagnostic; forensic; genetic analysis; ss.
XX
OS   Synthetic.
XX
PN   WO200004192-A1.
XX
PD   27-JAN-2000.
XX
PF   16-JUL-1999;   99WO-US016242.
XX
PR   17-JUL-1998;   98US-0093219P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Shafer DA;
XX
DR   WPI; 2000-182448/16.
XX
PT   New Gene-tag reporter for joining to nucleic acid probe, used, e.g. for
PT   mapping genes or mutational analysis.
XX
PS   Example 15; Page 70; 164pp; English.
XX
CC   This sequence is used in the construction of the gene-tag reporter of the
CC   invention. The invention relates to a gene-tag reporter for joining a
CC   probe, alone or with a second gene-tag reporter. The gene-tag reporter
CC   comprises a labelled double-stranded polynucleotide sequence having one
CC   or more linkers that comprise a single stranded sequence hybridisable to
CC   a complement but not to the target probe. Also included in the invention
CC   is a reporter array, comprising at least two gene-tag reporters linked
CC   end to end by hybridisable linkers. The reporters are used in various new
CC   methods for detecting and mapping genes; identifying mutations and
CC   variant nucleic acids, e.g. detecting rare mutations such as those in
CC   cancer cells or mutant viruses, but more generally in human diagnostics,
CC   forensics, genetic analysis, analysis of environmental samples or foods.
CC   The gene-tag reporters and associated probes, have a modular structure,
CC   allowing simple and inexpensive probe design, and are able to generate a
CC   distinctive signal, based on the mix and/or proportions of different
CC   signal components. Many targets can be analysed simultaneously, using
CC   many probes. Arrays of gene-tag reporters will provide signal
CC   amplification
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 27;  DB 3;  Length 27;
  Best Local Similarity  100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 12
AAS14192/c
ID   AAS14192 standard; DNA; 27 BP.
XX
AC   AAS14192;
XX
DT   18-DEC-2001  (first entry)
XX
DE   GeneTAG First-GREEN primer used in construction of probe sets.
XX
KW   WRAP-Probe; gene expression array; global amplification; RNA array; ss;
KW   tissue microarray; drug discovery assay; reporter binding site; forensic;
KW   diagnostic; genomic analysis; universal linker; PCR primer.
XX
OS   Synthetic.
XX
PN   WO200166802-A1.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US007508.
XX
PR   09-MAR-2000; 2000US-0187982P.
XX
PA   (GENE-) GENETAG TECHNOLOGY INC.
XX
PI   Shafer DA;
XX
DR   WPI; 2001-596845/67.
XX
PT   Novel probe sets with common universal linkers at one or both ends (WRAP
PT   probes) for gene expression arrays to provide global amplification of
PT   probe set and to provide common equivalent signaling regardless of
PT   length.
XX
PS   Example 3; Page 59; 97pp; English.
XX
CC   The invention relates to a probe set for gene expression arrays to
CC   provide common equivalent signalling per probe and global amplification
CC   of the set. The probe set has a pool of modified cDNA probes, each probe
CC   having a central target specific segment copied from a portion of a
CC   single mRNA transcript and a universal linker (a WRAP-Probe) located on
CC   one or both terminal ends. The universal linker has reporter binding
CC   sites to join common reporters to the probes and primer binding sites to
CC   copy and amplify the probe. The probes and reporters are useful in
CC   diagnostic or drug discovery assays for a wide range of biomedical
CC   samples, including detection of nucleic acids and gene expression
CC   profiles in human diagnostics, forensics and genomic analysis. The
CC   methods are useful for amplifying and identifying any unknown DNA
CC   fragment and also for improving sensitivity with tissue microarrays or
CC   RNA arrays. The methods improve the quantification of gene expression and
CC   allow highly improved detection of rare transcripts or very small
CC   samples. This sequence represents a GeneTAG First-GREEN primer used in
CC   the construction of probe sets
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 27;  DB 4;  Length 27;
  Best Local Similarity  100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 13
AAS14194/c
ID   AAS14194 standard; DNA; 27 BP.
```

```
XX
AC   AAS14194;
XX
DT   18-DEC-2001  (first entry)
XX
DE   First-GREEN ChipTAG primer used in construction of probe sets.
XX
KW   WRAP-Probe; gene expression array; global amplification; RNA array; ss;
KW   tissue microarray; drug discovery assay; reporter binding site; forensic;
KW   diagnostic; genomic analysis; universal linker; PCR primer.
XX
OS   Synthetic.
XX
PN   WO200166802-A1.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US007508.
XX
PR   09-MAR-2000; 2000US-0187982P.
XX
PA   (GENE-) GENETAG TECHNOLOGY INC.
XX
PI   Shafer DA;
XX
DR   WPI; 2001-596845/67.
XX
PT   Novel probe sets with common universal linkers at one or both ends (WRAP
PT   probes) for gene expression arrays to provide global amplification of
PT   probe set and to provide common equivalent signaling regardless of
PT   length.
XX
PS   Example 4; Page 60; 97pp; English.
XX
CC   The invention relates to a probe set for gene expression arrays to
CC   provide common equivalent signalling per probe and global amplification
CC   of the set. The probe set has a pool of modified cDNA probes, each probe
CC   having a central target specific segment copied from a portion of a
CC   single mRNA transcript and a universal linker (a WRAP-Probe) located on
CC   one or both terminal ends. The universal linker has reporter binding
CC   sites to join common reporters to the probes and primer binding sites to
CC   copy and amplify the probe. The probes and reporters are useful in
CC   diagnostic or drug discovery assays for a wide range of biomedical
CC   samples, including detection of nucleic acids and gene expression
CC   profiles in human diagnostics, forensics and genomic analysis. The
CC   methods are useful for amplifying and identifying any unknown DNA
CC   fragment and also for improving sensitivity with tissue microarrays or
CC   RNA arrays. The methods improve the quantification of gene expression and
CC   allow highly improved detection of rare transcripts or very small
CC   samples. This sequence represents a First-GREEN ChipTAG primer used in
CC   the construction of probe sets
XX
SQ   Sequence 27 BP; 6 A; 7 C; 9 G; 5 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db         27 GTAGCCTAGCTACCCCTAGGTCTAGGC 1


RESULT 14
AAS14183
ID   AAS14183 standard; DNA; 27 BP.
XX
AC   AAS14183;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Fragment #1 of PCR primer GR-SPC-F used in construction of probe sets.
XX
KW   WRAP-Probe; gene expression array; global amplification; RNA array; ss;
KW   tissue microarray; drug discovery assay; reporter binding site; forensic;
KW   diagnostic; genomic analysis; universal linker; PCR primer.
XX
OS   Synthetic.
XX
PN   WO200166802-A1.
XX
PD   13-SEP-2001.
XX
PF   09-MAR-2001; 2001WO-US007508.
XX
PR   09-MAR-2000; 2000US-0187982P.
XX
PA   (GENE-) GENETAG TECHNOLOGY INC.
XX
PI   Shafer DA;
XX
DR   WPI; 2001-596845/67.
XX
PT   Novel probe sets with common universal linkers at one or both ends (WRAP
PT   probes) for gene expression arrays to provide global amplification of
PT   probe set and to provide common equivalent signaling regardless of
PT   length.
XX
PS   Disclosure; Page 90; 97pp; English.
XX
CC   The invention relates to a probe set for gene expression arrays to
CC   provide common equivalent signalling per probe and global amplification
CC   of the set. The probe set has a pool of modified cDNA probes, each probe
CC   having a central target specific segment copied from a portion of a
CC   single mRNA transcript and a universal linker (a WRAP-Probe) located on
CC   one or both terminal ends. The universal linker has reporter binding
CC   sites to join common reporters to the probes and primer binding sites to
CC   copy and amplify the probe. The probes and reporters are useful in
CC   diagnostic or drug discovery assays for a wide range of biomedical
CC   samples, including detection of nucleic acids and gene expression
CC   profiles in human diagnostics, forensics and genomic analysis. The
CC   methods are useful for amplifying and identifying any unknown DNA
CC   fragment and also for improving sensitivity with tissue microarrays or
CC   RNA arrays. The methods improve the quantification of gene expression and
CC   allow highly improved detection of rare transcripts or very small
CC   samples. This sequence represents a fragment of a PCR primer used in the
CC   construction of probe sets
XX
SQ   Sequence 27 BP; 5 A; 9 C; 7 G; 6 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 27;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 0.0075;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||||||||||||||||||||
Db          1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27


RESULT 15
AAS14198
ID   AAS14198 standard; DNA; 27 BP.
XX
AC   AAS14198;
XX
DT   18-DEC-2001  (first entry)
XX
DE   First-GREEN random adapter part 2 used in construction of probe sets.
XX
KW   WRAP-Probe; gene expression array; global amplification; RNA array; ss;
KW   tissue microarray; drug discovery assay; reporter binding site; forensic;
KW   diagnostic; genomic analysis; universal linker.
XX
OS   Synthetic.
XX
PN   WO200166802-A1.
```

XX
PD 13-SEP-2001.
XX
PF 09-MAR-2001; 2001WO-US007508.
XX
PR 09-MAR-2000; 2000US-0187982P.
XX
PA (GENE-) GENETAG TECHNOLOGY INC.
XX
PI Shafer DA;
XX
DR WPI; 2001-596845/67.
XX
PT Novel probe sets with common universal linkers at one or both ends (WRAP
PT probes) for gene expression arrays to provide global amplification of
PT probe set and to provide common equivalent signaling regardless of
PT length.
XX
PS Example 7; Page 65; 97pp; English.
XX
CC The invention relates to a probe set for gene expression arrays to
CC provide common equivalent signalling per probe and global amplification
CC of the set. The probe set has a pool of modified cDNA probes, each probe
CC having a central target specific segment copied from a portion of a
CC single mRNA transcript and a universal linker (a WRAP-Probe) located on
CC one or both terminal ends. The universal linker has reporter binding
CC sites to join common reporters to the probes and primer binding sites to
CC copy and amplify the probe. The probes and reporters are useful in
CC diagnostic or drug discovery assays for a wide range of biomedical
CC samples, including detection of nucleic acids and gene expression
CC profiles in human diagnostics, forensics and genomic analysis. The
CC methods are useful for amplifying and identifying any unknown DNA
CC fragment and also for improving sensitivity with tissue microarrays or
CC RNA arrays. The methods improve the quantification of gene expression and
CC allow highly improved detection of rare transcripts or very small
CC samples. This sequence represents a random adapter fragment used in the
CC construction of probe sets
XX
SQ Sequence 27 BP; 5 A; 9 C; 7 G; 6 T; 0 U; 0 Other;

Query Match 100.0%; Score 27; DB 4; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.0075;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27
||||||||||||||||||||||||||||
Db 1 GTAGCCTAGCTACCCCTAGGTCTAGGC 27

Search completed: October 13, 2005, 19:12:37
Job time : 1123 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 13, 2005, 18:25:43 ; Search time 7556 Seconds
(without alignments)
136.016 Million cell updates/sec

Title: US-09-744-097A-76
Perfect score: 27
Sequence: 1 gtagcctagctacccctaggtctaggc 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		19.6	72.6	297	2	BB274268	BB274268 BB274268
2		19.6	72.6	623	1	AV325957	AV325957 AV325957
c 3		19.6	72.6	801	2	BE394439	BE394439 601311706
c 4		19.6	72.6	896	2	BE314562	BE314562 601151354
5		19.6	72.6	3378	3	AK032394	AK032394 Mus muscu
6		19.2	71.1	115	9	CG805069	CG805069 1118057A0
7		19.2	71.1	115	9	CG805158	CG805158 1118058C0
8		19.2	71.1	419	8	AZ636851	AZ636851 1M0495I22
9		19.2	71.1	453	9	CC684396	CC684396 OGUKC11TV
c 10		19.2	71.1	630	8	BZ631002	BZ631002 PUAAZ80TD
11		19.2	71.1	672	8	BZ630997	BZ630997 PUAAZ80TB
12		19.2	71.1	683	8	BZ633850	BZ633850 PUAAY40TB
c 13		19.2	71.1	705	8	BH837552	BH837552 LMCR10000
c 14		19.2	71.1	725	8	BZ633854	BZ633854 PUAAY40TD
c 15		19.2	71.1	745	9	CG351990	CG351990 OGVCK84TH
c 16		19.2	71.1	747	9	CC626350	CC626350 OGUJV06TH
17		19.2	71.1	828	9	CC686399	CC686399 OGULV40TV
c 18		19.2	71.1	839	9	CG306780	CG306780 OG1AT60TH
c 19		19.2	71.1	875	9	CG227499	CG227499 OG2BS48TV
c 20		19.2	71.1	902	9	CR222206	CR222206 Reverse s
c 21		19.2	71.1	951	8	CC007567	CC007567 PUGKB80TD
22		19	70.4	478	7	CO781043	CO781043 BL011B_D0
c 23		19	70.4	562	8	BZ306625	BZ306625 hx44g11.b
24		19	70.4	774	8	CC122296	CC122296 NDL.65A18
25		19	70.4	779	9	CC874542	CC874542 NDL.131O9
c 26		19	70.4	786	9	CG063671	CG063671 PUFTI87TD
27		19	70.4	864	9	CG180375	CG180375 PUFTJ86TB
c 28		19	70.4	865	9	CG180377	CG180377 PUFTJ86TD
29		18.8	69.6	535	8	AZ500954	AZ500954 1M0339N06
c 30		18.8	69.6	578	8	AZ492835	AZ492835 1M0327K11
31		18.8	69.6	671	9	CR270282	CR270282 Reverse s
32		18.8	69.6	675	9	CR273252	CR273252 Reverse s
c 33		18.8	69.6	692	9	AG293534	AG293534 Mus muscu
c 34		18.8	69.6	711	9	AG286334	AG286334 Mus muscu
35		18.8	69.6	720	8	BZ257884	BZ257884 CH230-479
c 36		18.8	69.6	721	9	AG567157	AG567157 Mus muscu
c 37		18.6	68.9	806	7	CK316863	CK316863 SB02020A1
38		18.6	68.9	837	4	BG211849	BG211849 RST31422
c 39		18.6	68.9	869	8	BZ785030	BZ785030 PUGDB13TD
c 40		18.6	68.9	880	8	BZ401569	BZ401569 OGACU96TC
41		18.6	68.9	954	8	CC295826	CC295826 CH261-73H
42		18.6	68.9	998	9	CG424886	CG424886 ZMMBBc004
43		18.6	68.9	1039	9	CC809843	CC809843 ZMMBBc049
c 44		18.2	67.4	406	8	AZ508009	AZ508009 1M0349L24
45		18.2	67.4	428	6	CB097308	CB097308 554225 MA

ALIGNMENTS

RESULT 1
BB274268
LOCUS BB274268 297 bp mRNA linear EST 07-JUL-2000
DEFINITION BB274268 RIKEN full-length enriched, 10 days neonate cortex Mus musculus cDNA clone A830087F07 3', mRNA sequence.
ACCESSION BB274268
VERSION BB274268.1 GI:8971289
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 297)
AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL Unpublished (2000)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .297
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A830087F07"
/tissue_type="cortex"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days neonate cortex"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match 72.6%; Score 19.6; DB 2; Length 297;
Best Local Similarity 84.6%; Pred. No. 2.9e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
|||||||||| ||||| || |||||
Db 243 TAGCCTAGCTTCCCCTTTGTTTAGGC 268

RESULT 2
AV325957
LOCUS AV325957 623 bp mRNA linear EST 24-OCT-2001
DEFINITION AV325957 RIKEN full-length enriched, adult male medulla oblongata Mus musculus cDNA clone 6330407C01 3', mRNA sequence.
ACCESSION AV325957
VERSION AV325957.2 GI:16395524
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 623)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT On Nov 11, 1999 this sequence version replaced gi:6366009.
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Location/Qualifiers
source 1. .623
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330407C01"
/sex="male"
/tissue_type="medulla oblongata"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male medulla oblongata"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match 72.6%; Score 19.6; DB 1; Length 623;
Best Local Similarity 84.6%; Pred. No. 3e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
|||||||||| ||||| || |||||
Db 569 TAGCCTAGCTTCCCCTTTGTTTAGGC 594

RESULT 3

```
BE394439/c
LOCUS       BE394439                 801 bp    mRNA    linear   EST 21-JUL-2000
DEFINITION  601311706F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3633168 5',
            mRNA sequence.
ACCESSION   BE394439
VERSION     BE394439.1  GI:9339804
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM325  row: l  column: 01
            High quality sequence stop: 390.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3633168"
                     /tissue_type="endometrium, adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_44"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies)."
ORIGIN

  Query Match             72.6%;  Score 19.6;  DB 2;  Length 801;
  Best Local Similarity   84.6%;  Pred. No. 3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||| | || |||||||| |||||
Db        154 TAGCCTTGTTATCCCTAGGTATAGGC 129


RESULT 4
BE314562/c
LOCUS       BE314562                 896 bp    mRNA    linear   EST 26-OCT-2000
DEFINITION  601151354F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3507589 5',
            mRNA sequence.
ACCESSION   BE314562
VERSION     BE314562.1  GI:9135847
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 896)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM187  row: k  column: 14
            High quality sequence start: 78
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1. .896
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3507589"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_19"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             72.6%;  Score 19.6;  DB 2;  Length 896;
  Best Local Similarity   84.6%;  Pred. No. 3e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||| | || |||||||| |||||
Db        235 TAGCCTTGTTATCCCTAGGTATAGGC 210


RESULT 5
AK032394
LOCUS       AK032394                 3378 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus adult male olfactory brain cDNA, RIKEN full-length
            enriched library, clone:6430537F04 product:unknown EST, full insert
            sequence.
ACCESSION   AK032394
VERSION     AK032394.1  GI:26082821
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
```

```
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3378)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .3378
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:6430537F04"
                     /db_xref="taxon:10090"
                     /clone="6430537F04"
                     /sex="male"
                     /tissue_type="olfactory brain"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1. .3378
                     /note="unknown EST (GB|AW494659, evidence: BLASTN, 99%,
                     match=458)"
ORIGIN

  Query Match             72.6%;  Score 19.6;  DB 3;  Length 3378;
  Best Local Similarity   84.6%;  Pred. No. 3.1e+02;
  Matches   22;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          2 TAGCCTAGCTACCCCTAGGTCTAGGC 27
              |||||||||| |||||  || |||||
Db        851 TAGCCTAGCTTCCCCTTTGTTTAGGC 876


RESULT 6
CG805069
LOCUS       CG805069                 115 bp    DNA     linear   GSS 10-NOV-2003
DEFINITION  1118057A09.x1 1118 - RescueMu Grid S Zea mays genomic, genomic
            survey sequence.
ACCESSION   CG805069
VERSION     CG805069.1  GI:38241110
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site so sequence was trimmed. Post-ligation
            sequence submitted separately.
            Plate: 1118057  row: 10
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .115
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="1118 - RescueMu Grid S"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid S was grown at San Diego in 2002. DNA was
                     extracted from leaf strips, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 9;  Length 115;
  Best Local Similarity   87.5%;  Pred. No. 4.3e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 GCCTAGCTACCCCTAGGTCTAGGC 27
              |||| ||||||||||| ||||||
Db         58 GCCTGGCTACCCCTAGCCCTAGGC 81


RESULT 7
CG805158
LOCUS       CG805158                 115 bp    DNA     linear   GSS 10-NOV-2003
DEFINITION  1118058C04.x1 1118 - RescueMu Grid S Zea mays genomic, genomic
            survey sequence.
ACCESSION   CG805158
VERSION     CG805158.1  GI:38241278
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 115)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
```

```
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site so sequence was trimmed. Post-ligation
            sequence submitted separately.
            Plate: 1118058  row: 10
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .115
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="1118 - RescueMu Grid S"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the web
                     site 'www.zmdb.iastate.edu' and follow the links for
                     'RescueMu.' Grid S was grown at San Diego in 2002. DNA was
                     extracted from leaf strips, double digested using BamHI
                     and BglII, and ligated to form circular plasmids. DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 9;  Length 115;
  Best Local Similarity   87.5%;  Pred. No. 4.3e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 GCCTAGCTACCCCTAGGTCTAGGC 27
              |||| |||||||||||  ||||||
Db         58 GCCTGGCTACCCCTAGCCCTAGGC 81


RESULT 8
AZ636851
LOCUS       AZ636851                 419 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  1M0495I22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0495I22 R, genomic survey sequence.
ACCESSION   AZ636851
VERSION     AZ636851.1  GI:11759041
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0495  row: I  column: 22
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0495I22"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 8;  Length 419;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AGCCTAGCTACCCCTAGGTCTAGG 26
              || |||||| ||||||||||| |||
Db         51 AGTCTAGCTTCCCCTAGGTCAAGG 74


RESULT 9
CC684396
LOCUS       CC684396                 453 bp    DNA     linear   GSS 19-JUN-2003
DEFINITION  OGUKC11TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0459B21,
            genomic survey sequence.
ACCESSION   CC684396
VERSION     CC684396.1  GI:32089172
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 453)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGUKC11TH
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
```

```
     source          1. .453
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0459B21"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 9;  Length 453;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTA 24
              |||||||||||||||||||  | |
Db         83 GTAGCCTAGCTACCCCTAGCACCA 106


RESULT 10
BZ631002/c
LOCUS       BZ631002                630 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  PUAAZ80TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa007M16,
            genomic survey sequence.
ACCESSION   BZ631002
VERSION     BZ631002.1  GI:28077996
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 630)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUAAZ80TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa007M16"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 8;  Length 630;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTA 24
              |||||||| | ||||||||||  ||
Db        472 GTAGCCTATCCACCCCTAGGTGTA 449


RESULT 11
BZ630997
LOCUS       BZ630997                672 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  PUAAZ80TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa007M16,
            genomic survey sequence.
ACCESSION   BZ630997
VERSION     BZ630997.1  GI:28077991
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 672)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUAAZ80TD
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa007M16"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 8;  Length 672;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTA 24
              |||||||| | ||||||||||  ||
Db        116 GTAGCCTATCCACCCCTAGGTGTA 139


RESULT 12
BZ633850
LOCUS       BZ633850                683 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  PUAAY40TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa007G07,
            genomic survey sequence.
ACCESSION   BZ633850
VERSION     BZ633850.1  GI:28080844
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUAAY40TD
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
```

```
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa007G07"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 8;  Length 683;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTA 24
              |||||||| | |||||||||| ||
Db        116 GTAGCCTATCCACCCCTAGGTGTA 139


RESULT 13
BH837552/c
LOCUS       BH837552                705 bp    DNA     linear   GSS 28-MAY-2002
DEFINITION  LMCR100001A03f Zea mays L. Zea mays genomic clone LMCR100001A03f,
            genomic survey sequence.
ACCESSION   BH837552
VERSION     BH837552.1  GI:21235430
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Kim,S.W., Yu,Y., Lee,M.C., Main,D. and Wing,R.A.
  TITLE     Methyl-filtration genomic sequence from maize
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total High Quality bases = 551
            Seq primer: TAATACGACTCACTATAGGG
            Class: shotgun
            High quality sequence start: 12
            High quality sequence stop: 690.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="LMCR100001A03f"
                     /tissue_type="Leaf"
                     /lab_host="DH10B"
                     /clone_lib="Zea mays L."
                     /note="Vector: pGEM-T easy; Site_1: Mcr BC;
                     Methyl-filtration library, Nuclei DNA was completely
                     digested with  Mcr BC, size fractionated and transformed
                     to E.Coli.DH10B."
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 8;  Length 705;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 GCCTAGCTACCCCTAGGTCTAGGC 27
              |||| ||||||||||  ||||||
Db        128 GCCTGGCTACCCCTAGCCCTAGGC 105


RESULT 14
BZ633854/c
LOCUS       BZ633854                725 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  PUAAY40TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa007G07,
            genomic survey sequence.
ACCESSION   BZ633854
VERSION     BZ633854.1  GI:28080848
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUAAY40TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa007G07"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 8;  Length 725;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 GTAGCCTAGCTACCCCTAGGTCTA 24
              |||||||| | |||||||||| ||
Db        567 GTAGCCTATCCACCCCTAGGTGTA 544


RESULT 15
CG351990/c
LOCUS       CG351990                745 bp    DNA     linear   GSS 26-AUG-2003
DEFINITION  OGVCK84TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0504N23,
            genomic survey sequence.
ACCESSION   CG351990
VERSION     CG351990.1  GI:34269256
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGVCK84TV
            Contact: Cathy Whitelaw
            TIGR
```

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

```
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0504N23"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             71.1%;  Score 19.2;  DB 9;  Length 745;
  Best Local Similarity   87.5%;  Pred. No. 4.5e+02;
  Matches   21;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          4 GCCTAGCTACCCCTAGGTCTAGGC 27
              |||| |||||||||||  ||||||
Db        389 GCCTGGCTACCCCTAGCCCTAGGC 366


Search completed: October 13, 2005, 21:43:33
Job time : 7564 secs
```